

Production et mise en ligne d'un dictionnaire électronique du wolof

Chérif MBODJ¹ Chantal ENGUEHARD²

(1) CLAD1, Université Cheikh Anta Diop de Dakar, Sénégal

(2) LINA, Université de Nantes, France

cherif.mbodj@ucad.edu.sn, chantal.enguehard@univ-nantes.fr

RÉSUMÉ

Cet article présente l'état actuel des résultats du projet de mise en ligne d'un dictionnaire électronique du wolof, soutenu par l'Université de Nantes, auquel travaillent, Chantal Enguehard, une informaticienne de LINA et Chérif Mbodj, un linguiste du CLAD. Ce projet s'inscrit dans le cadre global du projet Dictionnaires Langue Africaine-Français (DiLAF), réalisé à l'Université Nantes et financé à l'origine par l'Organisation Internationale de la Francophonie. Nous avons mis en œuvre la méthodologie DiLAF, dont les objectifs sont, d'une part, de produire des dictionnaires au format XML1 compatibles avec le standard international Lexical Markup Framework (LMF) de balisage de dictionnaires et, d'autre part, de générer des dictionnaires HTML2 accessibles en ligne (Enguehard et Mangeot 2015). La méthodologie DiLAF vise à pallier l'insuffisante présence des langues peu dotées, comme les langues africaines, sur la Toile.

ABSTRACT

A Wolof Dictionary online

This paper deals with the present state of the outcome of the project to put on-line an electronic Wolof dictionary, supported by the University of Nantes, and on which are working, Chantal Enguehard, a computer scientist at LINA and Chérif Mbodj, a linguist from the Center for Applied Linguistics of Dakar - CLAD.

That project is in keeping with the general pattern of the DiLAF project set up by the Nantes University and initially funded by the International Organization of French-Speaking World. We have implemented the DiLAF methodology the objectives of which are, on the one hand, to produce dictionaries of XML format compatible with the Lexical Markup Framework standard (LMF) of dictionary markup, and on the other hand, to generate HTML dictionaries accessible on-line (Enguehard and Mangeot, 2015). Besides, the DiLAF methodology aims at mitigating the weak presence of less-resourced languages such as the African languages on the Web.

TËNK

Baatukaay wolof ci enternet

Bataxal bi dafay wone tolluwaayu mujjantalu mébétum samp baatukaayu wolof ci këru enternet, bi nga xam né Daara ju mag ju Nànt moy ciy joxe ndimbalam te Santaal Angëhaar, gëstukatu xam-xamu xarala ci LINA ak Sërif Mbóoj, gëstukatu kàllaama ci CLAD di ciy liggéey. Mébét mii mu ngi bindu ci mébétu Baatukaayu Làkku Afrig-Franse (BaLAF)1bi Daara ju mag ju Nànt di doxal te Kuréelu Réew yiy Làkk Faranse jàppale ko ci. Mébét mii mu ngiy jëfandikoo doxalinu BaLAF bi nga xam ne li ko taxa jóg mooy, benn, móol ay baatukaay yu bennalinu ay denc XML yu méngook gindikaayu réew yepp ci raññeekayu baatukaay, LMF, ñaarel ba, moy móol ay baatukaay HTLM,

yuuy yombalu jëriñoo ci enternet (Angëhaar ak Mànjó 2015). BaLAF li ko soxal moy fexe ba dàgg li waral làkk yi rafle téere, mel ni làkku Afrig, seen teeway ci enternet amul solo.

MOTS-CLÉS : Traitement automatique des langues, outils informatiques, langues africaines, lexicologie.

KEYWORDS: Language automatic processing, computing tools, African languages, lexicology, terminology.

BAATU GINDIKAAY: Liggéeyu làkk bu dox biiru, juntuukaayu enternet, làkku Afrig, mbirum sàqum-baat, mbiru sàqum-baatu xeetu xam-xam.

Introduction

Le LINA et le CLAD travaillent actuellement à la mise en ligne d'un dictionnaire électronique du wolof.

L'objectif de cette contribution est de présenter les premiers résultats de ce projet, soutenu par l'Université de Nantes, et de les soumettre à l'appréciation des linguistes et des spécialistes du traitement automatique des langues africaines participant à TALAF 2016.

Nous procéderons, d'abord, à une présentation du projet, ensuite, à celle du dictionnaire d'origine et, enfin, à celle des transformations des ressources lexicales recueillies, suite à l'application de la méthodologie DILAF (Dictionnaires Langues Africaines - Français).

1 Présentation du projet

1.1 Objectif global

L'objectif global du projet est d'assurer la présence de langues africaines d'Afrique de l'Ouest sur la Toile, notamment le wolof (parlé au Sénégal par plus de 80% de la population, en Mauritanie et en Gambie) et de constituer des ressources utilisables en Traitement Automatique des Langues (TAL). Il s'agit d'encourager la production de pages web bilingues et contribuer ainsi à la réduction de la fracture numérique entre langues bien dotées et langues africaines.

1.2 Objectifs spécifiques

Le projet vise trois objectifs spécifiques.

- Il s'agit de formaliser des dictionnaires relatifs à des langues africaines en versions XML utilisables pour le TAL. Les dictionnaires sont transformés en suivant la méthodologie DiLAF (projet réalisé au LINA, financé pour l'Organisation Internationale de la Francophonie (O.I.F), il a donné lieu à plusieurs publications figurant en bibliographie). Les fichiers XML sont téléchargeables et peuvent être utilisés dans le respect de la licence¹ sous laquelle ils sont publiés.
- Les dictionnaires une fois formalisés peuvent être affichés sur la Toile au bénéfice des wolophones.

¹ Creative Attribution-NonCommercial-ShareAlike 2.0 Generic (CC BY-NC-SA 2.0)

- Ce travail mené en collaboration est l'occasion de transferts de connaissances en informatique vers des linguistes. Il s'agit d'acquérir et d'utiliser des savoirs concernant le codage de caractères, le balisage et l'usage des expressions rationnelles. L'usage des langues africaines dans des contextes scientifiques ou techniques est également encouragé par la création de nom de balises (cf. 3.2.3)

Cet article présente les étapes franchies en ce qui concerne un dictionnaire wolof pour établir une première version strictement limitée à une formalisation des informations présentes dans les fichiers que nous avons collectés. Les étapes de correction et d'enrichissement seront envisagées à l'avenir.

2 Ressources lexicales d'origine

Des équipes du CLAD et de l'IFAN² ont produit un fichier de ressources lexicales en compilant deux ouvrages respectivement intitulés :

1. *Lexique wolof-français*, en 4 volumes, (vol. 1 et 2 : 1977, vol. 3 : 1979, vol. 4 : 1981) ;
2. *Dictionnaire wolof - français / anglais* en 1 volume, une adaptation des 4 volumes précités.

Ces ressources ont été choisies car elles sont la propriété du CLAD. d'autres ressources (lexique ou dictionnaire) pourraient enrichir ce premier corpus dans la mesure où les auteurs seraient d'accord pour les partager et en fourniraient les fichiers sources.

2.1 Lexique wolof-français, en 4 volumes

Dans l'introduction du volume 1 (CLAD 1977 : 1) de ce lexique, il est précisé que :

"Le Wolof Fondamental dont l'élaboration a été entreprise en 1966 par le Centre de Linguistique Appliquée de Dakar, sur l'initiative de Jean Maurice Calvet et grâce à une subvention de la Fondation Ford, doit permettre de dégager les 6.000 mots les plus fréquents de la langue à partir d'un corpus de 500.000 mots traité mécanographiquement à Besançon³.

Les premiers résultats de cette vaste opération ont permis d'établir, dans un premier temps, un lexique Wolof-Français de 3.000 mots ; les mots retenus sont ceux que le dépouillement d'un sous-corpus de 200.000 mots fait apparaître comme étant les plus fréquents. Ce lexique est le fruit d'un travail collectif au sein d'une Commission à laquelle ont participé : Oumar Ben Khatab Dia, Aziz Diaw, El Hadj Cheick Gassama, (décédé), Mamadou Guèye, Saliou Kandji, Souleymane Ndiaye, Assane Sylla, Le Paster Church.

Oumar Marone, Papa Séga Name et Rawane Mbaye ont apporté leur précieux concours.

La mise au point linguistique a été assurée par Arame Diop Fal, Rosine Santos, Oumar Ben Khatab Dia, Jean Doneux et Mamadou Guèye. C'est grâce au soutien précieux de Pierre Dumont, Directeur du CLAD, que cette publication a été réalisée."

2.2 Dictionnaire wolof - anglais en 1 volume

Pour la Section d'anglais du CLAD, composée de Mary Saad, Cheikh Hamidou Kane et d'Ousseynou Ndiaye, il était clairement apparu que la traduction du lexique français-wolof était

² Institut fondamental d'Afrique noire,

³ Voir a) « Langage use and social change » edited by W. H. Whiteley, IAI Oxford University Press. 1971 (p. 274-287).

b) Les langues nationales et l'Enseignement : le projet wolof par Arame Diop Fal, in Recherche, Pédagogie et Culture n° 16, mars-avril 1975, p. 58-60.

devenu un impératif pour, d'une part, aider les anglophones à mieux maîtriser la langue wolof, et, d'autre part, pour faciliter l'apprentissage de l'anglais aux wolophones.

En réalité, la Section d'anglais a procédé à l'enrichissement des données existantes en y ajoutant des proverbes ou des expressions susceptibles de prendre en compte certains aspects de la culture anglaise.

La transcription phonétique a été revue par les linguistes de la Section des langues nationales du CLAD : Souleymane Faye, Chérif Mbodj et Ndiassé Thiam.

2.3 Structuration d'un article⁴

Chacun des articles du dictionnaire d'origine peut être décrit comme une séquence de neuf positions, dont certaines peuvent rester vacantes. Ainsi, chaque entrée wolof est suivie de :

- sa transcription phonétique ;
- sa catégorie grammaticale
- l'indication de son origine si elle est un emprunt ;
- ses valences si l'entrée est un verbe ;
- un renvoi à une variante potentielle ;
- sa spécification, le cas échéant, comme appartenant à un domaine scientifique ou technique ;
- son équivalent ou sa définition en français ;
- un exemple illustratif en wolof et sa traduction en français.

En cas de polysémie, les différents signifiés de l'entrée sont numérotés et traités dans le même article.

Exemples (cités exactement comme ils apparaissent dans les dictionnaires originaux) :

abal [ʌbʌ] *v.tr.* prêter - Abal ma sa tééré : *prête-moi ton livre.*

alxuraan j- [ʌlxura:n] *n.* *O.a* le Coran - Jàng naa alxuraan : *J'ai appris le Coran.*

loxo [lɔxɔ] *n.* 1. *anat.* bras, main - Loxo ndey jooram a damm : *C'est son bras droit qui est cassé.*
Prov. Dina mätt ag du mätt, bul jox sa loxo : *Dans le doute, on s'abstient.*

3 Transformation des ressources lexicales

3.1 Le projet DiLAF

Le projet DiLAF , en cours à LINA, et auquel s'ajoute le dictionnaire wolof, vise à convertir des dictionnaires éditoriaux bilingues (bambara, haoussa, kanouri, tamajaq, songhai-zarma) - français en un format XML permettant leur pérennisation et leur partage.

3.2 Méthodologie DiLAF

La méthodologie DiLAF comporte sept étapes :

- 1 : Prétraitements
- 2 : Conversion d'un fichier vers Unicode
- 3 : Choix des noms des éléments XML, structuration
- 4 : Balisage à l'aide d'expressions rationnelles

⁴ Partie du dictionnaire regroupant toutes les informations sur une même entrée.

- 5 : Contrôle des listes fermées
- 6 : Licence
- 7 : Documentation

Lors de nos premiers travaux, nous avons franchi les étapes détaillées ci-dessous.

3.2.1 Démarches préalables : Recueil des sources lexicales

Les dictionnaires précités ont été produits par le CLAD qui en détient donc les fichiers. Ces fichiers au format ".doc" ont été rassemblés en un seul fichier et la mise en page a été ôtée. Ce fichier sera nommé "le Dictionnaire" dans la suite de cet article.

3.2.2 Conversion d'un fichier vers Unicode

Le codage de l'alphabet selon Unicode a permis de produire l'alphabet et le tableau des codes Unicode mentionnés ci-dessous.

3.2.2.1 L'alphabet wolof

L'alphabet wolof est défini au Sénégal par le décret n° 71-566 du 21 mai 1971 relatif à la transcription des langues nationales. République du Sénégal, 1971.

Le décret n° 2005-992 du 21 octobre 2005 relatif à l'orthographe et à la séparation des mots en wolof est le décret actuellement en vigueur au Sénégal, concernant le wolof.

L'alphabet wolof comprend 63 symboles dont 32 digraphes :

Lettres minuscules :

a – à – ā – aa – āa – b – bb – c – cc – d – dd – e – ee – é – ée – ë – ëe – f – g – gg – i – ii – j – jj – k – kk – l – ll – m – mm – mb – mp – n – nn – nc – nd – ng – nj – nk – nq – nt – ñ – ññ – ŋ – ŋŋ – o – oo – ó – óo – p – pp – q – r – rr – s – t – tt – u – uu – w – ww – x – y – yy

Lettres majuscules à l'initiale d'un mot :

A – À – Ã – Aa – Æa – B – C – D – E – Ee – É – Ée – Ë – Ëe – F – G – I – Ii – J – K – L – M – Mb – Mp – N – Nc – Nd – Ng – Nj – Nk – Ñ – Ŋ – O – Oo – Ó – Óo – P – R – S – T – U – Uu – W – X – Y

3.2.2.2 Codes Unicode⁵

Les codes choisis pour représenter l'alphabet wolof sont présentés en annexe.

3.2.2.3 Conversion des caractères spéciaux

Les caractères spéciaux figurant dans le Dictionnaire étaient à l'origine écrits à l'aide d'une police "bricolée" dans laquelle le glyphe de certains caractères a été redessiné (Enguehard 2009). Repérer ces caractères et les convertir nécessite la collaboration entre linguiste et informaticien. Il peut s'agir de caractères utilisés pour écrire le wolof ou encore la phonétique.

Une fois repérés, ces caractères ont été convertis afin de n'obtenir que des caractères codés conformément à Unicode.

Voici quelques-uns de ces caractères convertis :

⁵ <http://www.unicode.org/charts/>

caractère original	caractère affiché
ɓ	ə
④	ʌ
☐	ʌ
⇒	j
⇒	ɲ

caractère original	caractère affiché
⑦	e
∇	ɔ
€	ε
α	ŋ
∇	ŋ

3.2.3 Éléments XML wolof

Les noms des éléments et des attributs XML utilisés pour baliser le Dictionnaire wolof ont fait l'objet d'une recherche terminologique fondée sur une approche culturelle (Diki-Kidiri 2008) qui a permis de les exprimer en wolof. Cette recherche devra être approfondie et affinée par une meilleure compréhension des notions informatiques.

La traduction des noms des balises en wolof devrait aider les locuteurs de cette langue à établir leur propre perception du traitement automatique des langues et à s'approprier les nouveaux savoirs et savoir-faire qui arrivent ainsi dans la société wolof.

Élément wolof	Équivalent français
bind	forme
bokk_tekki	synonyme
cosaan	origine
dog	article
faramfâce	explication
firi	traduction (d'un proverbe)
guddalinu_méngale	expression
léebu	proverbe
maas	classe
mbirmi	vedette
misaal	exemple
moomeelu_làkk	idiome
nettalin	version
nettalinu_tekki	version littérale
safaan	antonyme
sayu_tekki	bloc sémantique
tëggin	phonétique
tekki	équivalent
tool	domaine
xeet	catégorie grammaticale
xoolal	confer (cf)

weneen_waxin	variante
--------------	----------

Attribut wolof	Équivalent français
gàttal	abréviation
jëfandiko	emploi
làkk	langue
nii_mu_warloo	usage
waxin	style (niveau de langue)
jokkukaay	adresse (cible)
màndarga	identifiant

Les noms de langues ont été notés conformément au standard ISO 639-1 : "en" pour l'anglais, "fr" pour le français et "wo" pour le wolof.

3.2.4 Balisage à l'aide d'expressions rationnelles

Il s'agit de structurer l'entièreté du Dictionnaire à l'aide des balises précédemment définies. La transformation du Dictionnaire est réalisée par l'application judicieuse d'expressions régulières exploitant la mise en forme initiale (gras, italique, etc.). Cette approche permet un gain de temps considérable par rapport à la pose de balises une à une.

Cette étape exige une collaboration étroite entre linguiste et informaticien. D'une part, il est souvent impossible à l'informaticien non wolophone d'identifier les informations de chaque article. D'autre part, cette étape nécessite une utilisation experte des expressions régulières que le linguiste ne détient pas.

Lors de cette étape, il peut arriver que certains articles soient réunis car ils partagent le même groupe d'informations : vedette, classe, phonétique et catégorie grammaticale. Les blocs sémantiques restent, en revanche, différenciés.

3.2.4.1 Structuration d'un article

Les neuf (09) positions mentionnées sous 2.3. peuvent être considérées comme neuf (09) champs. Ceux-ci ont été requalifiés en balises de la façon suivante :

niveau 1	dog/màndarga	« article/identifiant »
niveau 2	mbirmi	« vedette »
niveau 3	tëggìn	« phonétique »
niveau 4	xeet/gàttal	« catégorie grammaticale /abréviation »
niveau 5	sayu_tekki/màndarga	« bloc sémantique/identifiant »
niveau 6	tekki	« équivalent »
niveau 7	nettalin/làkk="en"	« version/langue anglaise »
niveau 7	nettalin/làkk="fr"	« version/langue française »
niveau 8	Misaal	« Exemple »
niveau 9	nettalin/làkk="wo"	« version/langue wolof »
niveau 9	nettalin/làkk="en"	« version/langue anglaise »

niveau 9 nettalin/làkk="fr" « version/langue française »

3.2.4.2 Illustration des balises

Exemple 1 : Abal

```
<dog màndarga="abal1">
  <mbirmi>abal</mbirmi>
  <tëggin>[AbAl]</tëggin>
  <xeet gàttal="v.tr.">verbe transitif</xeet>
  <sayu_tekki màndarga="abal1_1">
    <tekki>
      <nettalin làkk="en"></nettalin>
      <nettalin làkk="fr">prêter</nettalin>
    </tekki>
    <misaal>
      <nettali làkk="wo">Abal ma sa tééré.</nettalin>
      <nettalin làkk="en"></nettalin>
      <nettali làkk="fr">prête-moi ton livre.</nettalin>
    </misaal>
  </sayu_tekki>
</dog>
```

Exemple 2 : alxuraan j-

```
<dog màndarga="alxuraan1">
  <mbirmi>alxuraan</mbirmi>
  <maas>j-</maas>
  <tëggin>[Alxura:n]</tëggin>
  <xeet gàttal="n.">nom</xeet>
  <cosaan>arabe</cosaan>
  <sayu_tekki màndarga="alxuraan1_1">
    <tekki>
      <nettalin làkk="en"></nettali>
      <nettalin làkk="fr">le Coran</nettalin>
    </tekki>
    <misaal>
      <nettalin làkk="wo">Jàng naa alxuraan.</nettalin>
      <nettalin làkk="en"></nettalin>
      <nettalin làkk="fr">J'ai appris le Coran.</nettalin>
    </misaal>
  </sayu_tekki>
</dog>
```

Exemple 3 : loxo b-

```
<dog màndarga="loxo1">
  <mbirmi>loxo</mbirmi>
  <maas>b-</maas>
  <tëggin>[lɔxɔ]</tëggin>
  <xeet gàttal="n.">nom</xeet>
  <sayu_tekki num="1" màndarga="loxo1_1">
    <tool>
      <nettalin làkk="fr">anatomie</nettalin>
    </tool>
  <tekki>
```



```

    <nettalin làkk="fr">avant_bras, main</nettalin>
  </tekki>
  <misaal>
    <nettalin làkk="wo">Loxo ndey jooram a damm.</nettalin>
    <nettalin làkk="fr">C'est son bras droit qui est cassé.</nettalin>
  </misaal>
  <léebu>
    <firi làkk="fr">
      <nettalinu_tekki>dans le doute entre être mordu ou pas, ne donne pas ta
      main.</nettalinu_tekki>
      <faramfacce>Dans le doute, on s'abstient.</faramfacce>
    </firi>
  </léebu>
</sayu_tekki>
</dog>

```

3.2.5 Documentation

La documentation associée au Dictionnaire a été réalisée :

- présentation de l'alphabet wolof
- présentation des éléments XML utilisés pour baliser le Dictionnaire
- présentation du Dictionnaire lui-même

Cette documentation sera accessible sur le site web DiLAF.

3.3 Affichage d'une page de consultation du Dictionnaire

Voici l'affichage des articles précédents :

abal [Abal]

Catégorie grammaticale : verbe transitif

- prêter

Exemple : Abal ma sa tééré. - prête-moi ton livre.

alxuraan (j-) [Alxura:n]

Catégorie grammaticale : nom

Origine : arabe

- le Coran

Exemple : Jàng naa alxuraan. - J'ai appris le Coran.

loxo [lɔxɔ]

Catégorie grammaticale : nom

- avant_bras, main

Domaine : anatomie

Exemple : Loxo ndey jooram a damm. - C'est son bras droit qui est cassé.

Proverbe : Dina mätt ag du mätt, bul jox sa loxo

litt : Dans le doute entre être mordu ou pas, ne donne pas ta main. - Dans le doute, on s'abstient.

4 Perspectives

Le Dictionnaire mis en ligne a été entièrement balisé. Cependant, il ne s'agit que d'une première version susceptible d'être améliorée. Plusieurs chantiers seront mis en œuvre en fonction de la force de travail que nous pourrions mobilier :

- Correction des erreurs au fur et à mesure de leur détection.
- Complétion des champs laissés vides étant donné l'absence de renseignements ou d'informations crédibles. La saisie de données complémentaires est en cours au CLAD.
- Evolution du dictionnaire bilingue wolof-français vers un dictionnaire trilingue wolof-français-anglais. Les balises mentionnées sous 3.2.4.1. ont déjà prévu cette possibilité.
- Ajout d'informations quant à la flexion et la dérivation afin que le Dictionnaire puisse être utilisé pour des applications de Traitement Automatique de Langues.

Conclusion

Le projet de mise en ligne du Dictionnaire wolof illustre comment la méthodologie DILAF peut aider à promouvoir davantage les langues africaines, pour ne citer que celles-là, en capitalisant sur les ressources existantes, en les enrichissant et en facilitant leur exploitation dans le but de contribuer ainsi à l'émergence de ces langues sur la Toile.

Les résultats que nous avons présentés ont été obtenus en un mois par une équipe de deux personnes : un linguiste et un informaticien. Cette expérience nous a permis de mesurer à quel point la collaboration entre les deux disciplines s'avère indispensable pour résoudre les questions apparues au fur et à mesure des traitements, qu'il s'agisse des questions d'encodage des caractères, de délimitation des informations, ou encore de structuration des articles.

Le dictionnaire sera prochainement mis en ligne sur le site web dilaf.org, accompagné de sa documentation. À notre connaissance, il s'agira du seul dictionnaire électronique wolof de cette ampleur, correctement documenté, et consultable sur la Toile.

Il rejoindra les dictionnaires déjà en ligne en bambara, haoussa, kanouri, tamajaq et zarma.

Bibliographie sélective

CLAD-IFAN. (1977). Lexique wolof-français, Tome 1 (lettres A-K), Dakar, CLAD.

DIKI-KIDIRI, M. (dir.) (2008). Le vocabulaire scientifique dans les langues africaines. Pour une approche culturelle de la terminologie. Paris, Karthala.

ENGUEHARD, C. (2009). Les langues d'Afrique de l'Ouest : de l'imprimante au traitement automatique des langues, Sciences et Techniques du Langage, 6, p.29-50. (ISSN 0850-3923).

ENGUEHARD, C., KANÉ, S., MANGEOT, M., Modi I., SANOGO, M. L. (2011). Vers l'informatisation de quelques langues d'Afrique de l'Ouest, 4ème atelier international sur l'amazighe et les TICs 'Les ressources langagières: construction et exploitation' ntic2011. Rabat, Maroc. 24 et 25 février 2011.

ENGUEHARD, C., KANÉ, S., MANGEOT, M., MODI, I., SANOGO, M.L. (juin 2012). Vers l'informatisation de quelques langues d'Afrique de l'Ouest, JEP-TALN-RECITAL 2012, Atelier TALAF 2012: Traitement Automatique des Langues Africaines, 27-40.

ENGUEHARD, C., MANGEOT, M. (2013). LMF for a selection of African Languages. Chapter 7. in Gil Francopoulo (dir.), LMF: Lexical Markup Framework, theory and practice. Ed., Hermès science. Paris, France. 17 p.

ENGUEHARD, C., MANGEOT, M. (2015). DILAF : des dictionnaires africains en ligne et une méthodologie. Actes du colloque 'Francophonie et langues nationales. UCAD, 21-22 novembre 2014, Dakar, Presses universitaires de Dakar, CLAD. 349-357.

GOVERNEMENT DU SÉNÉGAL (2006). Décret n° 2005-992 du 21 octobre 2005 relatif à l'orthographe et à la séparation des mots en wolof, J.O n° 6280 du 27 mai 2006, Primature, Dakar.

MANGEOT, M., ENGUEHARD, C. (2013). Des dictionnaires éditoriaux aux représentations XML standardisées. Chapter 8. in Gala, Núria et Michael Zock (dir.), Ressources Lexicales: Contenu, construction, utilisation, évaluation. 255–290.

MANGEOT, M., ENGUEHARD, C. (Septembre 2011). Informatisation de dictionnaires langues africaines-français. Actes des journées LTT 2011, Villetaneuse.

MBODJ, C. (2015). Terminologie, langues nationales et développement. Actes du colloque 'Francophonie et langues nationales, UCAD, 21-22 novembre 2014, Dakar, Presses universitaires de Dakar/ CLAD. 245-258.

ANNEXE

Symbole	Code Unicode	Symbole	Code Unicode
a	U+0061	A	U+0041
à	U+00E0	À	U+00C0
ã	U+00E3	Ã	U+00C3
b	U+0062	B	U+0042
bb	U+0062 U+0062		
c	U+0063	C	U+0043
cc	U+0063 U+0063		

d	U+0064	D	U+0044
dd	U+0064 U+0064		
e	U+0065	E	U+0045
ee	U+0065 U+0065	Ee	U+0045 U+0065
é	U+00E9	É	U+00C9
ée	U+00E9 U+00E9	Ée	U+00C9 U+0045
ë	U+00EB	Ë	U+00CB
ëe	U+00EB U+0065	Ëe	U+00CB U+0065
f	U+0066	F	U+0046
g	U+0067	G	U+0047
gg	U+0067 U+0067		
i	U+0069	I	U+0049
ii	U+0069 U+0069	Ii	U+0049 U+0069
j	U+006A	J	U+004A
jj	U+006A U+006A		
k	U+006B	K	U+004B
kk	U+006B U+006B		
l	U+006C	L	U+004C
ll	U+006C U+006C		
m	U+006D	M	U+004D
mm	U+006D U+006D		
mb	U+006D U+0062	Mb	U+004D U+0062
mp	U+006D U+0070	Mp	U+004D U+0070
n	U+006E	N	U+004E
nn	U+006E U+006E		
nc	U+006E U+0063	Nc	U+004E U+0063
nd	U+006E U+0064	Nd	U+004E U+0064
ng	U+006E U+0067	Ng	U+004E U+0067
nj	U+006E U+006A	Nj	U+004E U+006A
nk	U+006E U+006B	Nk	U+004E U+006B
nq	U+006E U+0071		
nt	U+006E U+0074		
ñ	U+00F1	Ñ	U+00D1
ññ	U+00F1 U+00F1		
ŋ	U+014B	Ŋ	U+014A
ŋŋ	U+014B U+014B		

o	U+006F	O	U+004F
oo	U+006F U+006F	Oo	U+004F U+006F
ó	U+00F3	Ó	U+00D3
óo	U+00F3 U+004F	Óo	U+00D3 U+006F
p	U+0070	P	U+0050
pp	U+0070 U+0070		
q	U+0071		
r	U+0072	R	U+0052
rr	U+0072 U+0072		
s	U+0073	S	U+0053
t	U+0074	T	U+0054
tt	U+0074 U+0074		
u	U+0075	U	U+0055
uu	U+0075 U+0075	Uu	U+0055 U+0075
w	U+0077	W	U+0057
ww	U+0077 U+0077		
x	U+0078	X	U+0058
y	U+0079	Y	U+0059
yy	U+0079 U+0079		

Notation du degré d'aperture

L'**accent aigu** sert à indiquer que la voyelle concernée est plus fermée que sa correspondante sans diacritique : ó / o, é / e. En revanche, l'**accent grave** indique que la voyelle est plus ouverte: à / a.